Analyzing Trends in Startup using SQL

Chapter 1

INTRODUCTION

In today's dynamic and ever-evolving business landscape, startups have emerged as significant players, driving innovation, disrupting traditional industries, and shaping the future of entrepreneurship. Understanding the trends within the startup ecosystem is of paramount importance, as it provides valuable insights into the strategies, challenges, and opportunities that these fledgling companies face. To unravel the intricate web of startup trends, we turn to the power of data analysis. SQL (Structured Query Language), a versatile tool for data management, offers us the means to explore and extract valuable information from startup databases, enabling us to draw informed conclusions.

Introduction to SQL:

SQL, or Structured Query Language, is a powerful and standardized programming language used for managing and manipulating relational databases. Developed in the 1970s, SQL has become the factory standard for working with data stored in a structured format, and it plays a crucial role in modern data management.

SQL is designed to perform a wide range of tasks related to database management, including:

1. Data Retrieval: SQL allows users to retrieve data from a database by using queries to specify what information they need. This is done using the SELECT statement.
2. Data Manipulation: SQL enables users to insert, update, and delete records in a database. The INSERT, UPDATE, and DELETE statements are used for these tasks.
3. Data Definition: SQL is used to define the structure of a database, including creating tables, defining their structure, setting constraints, and establishing relationships between tables.
4. Data Control: SQL provides mechanisms for controlling access to data, ensuring security, and maintaining data integrity through permission management.

Relational database management systems use structured query language (SQL) to store and manage data. The system stores multiple database tables that relate to each other. MS SQL Server, MySQL, or MS Access are examples of relational database management systems.

Chapter 2

Overview

Problem Statement :

It’s your first day as a Data Analyst. Your first task is to analyze the rising trends in the startup world, your boss emailed you a file that contains a Dataset called startups. It is a portfolio of some of the biggest names in the industry.

Role :

My role was to write queries to analyze and extract meaningful insights about these companies.

Approach :

I utilized SQLite for conducting various Queries. The dataset used for this project is provided by an online platform known as Codecademy. There is 1 file named project.SQLite that contains a table called startups. The startup's table contains the following columns:

name: Name of Startups

location: Location of startups

category: The category of startup

employees: Number of employees

raised: Funding raised by the startups

valuation: The valuation of the startup

founded: The year of foundation of the startup

stage: The current Stage of the startup

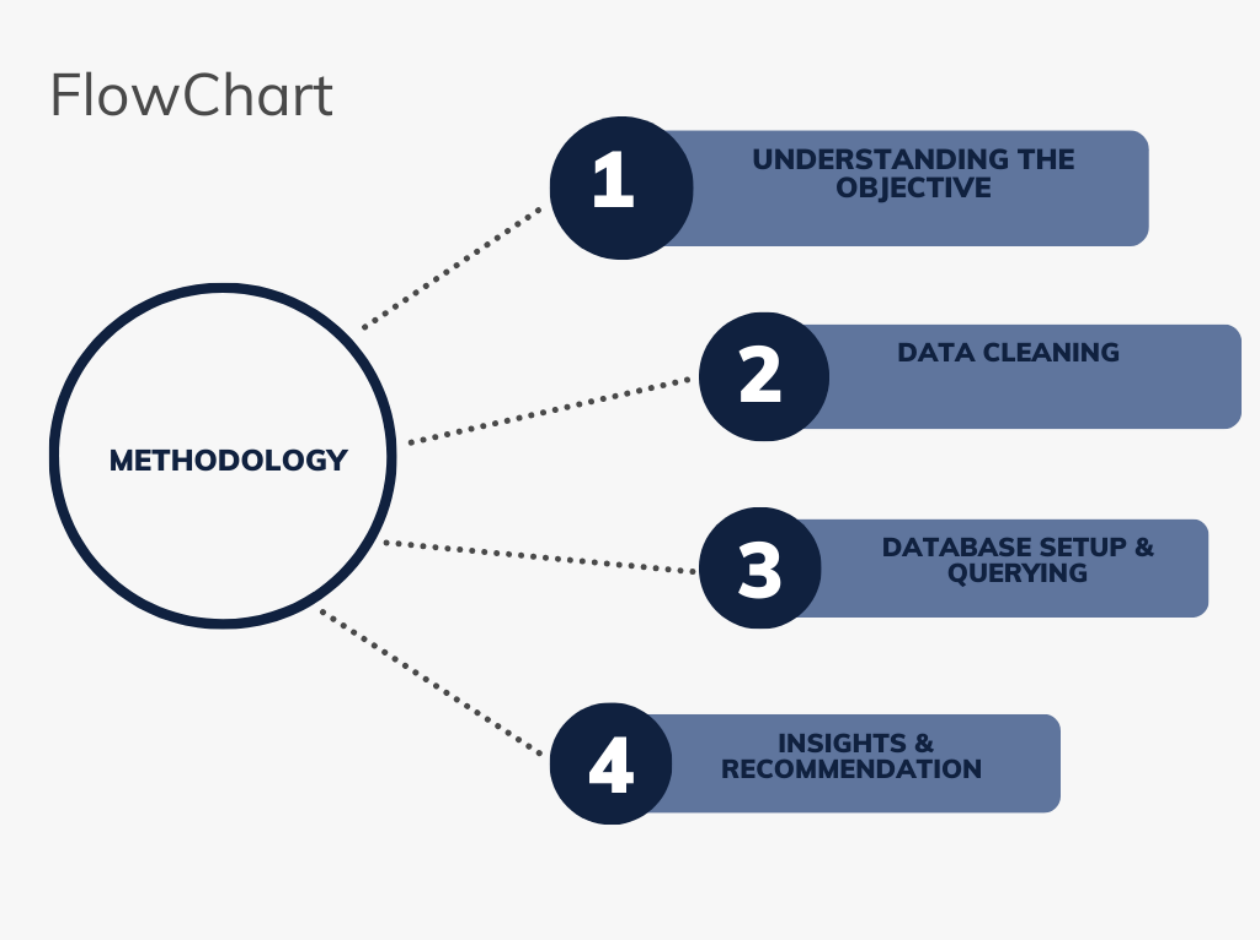
CEO: CEO of the startup

info: information about the startups

METHODOLOGY

Approach: I utilized SQLite to provide some useful insights.Wrote SQL queries to extract relevant data based on questions and demographics. The approach outlined in the methodology is structured and systematic, providing a step-by-step guide for conducting data analysis using SQL. It begins by familiarizing the user with the dataset and its structure. Subsequently, it addresses specific analytical tasks, starting with basic data retrieval and progressing to more complex calculations, filtering, and categorization.

The approach emphasizes clarity and comprehensibility, making it accessible to users with varying levels of SQL expertise. It ensures that data analysis is conducted methodically and organized, facilitating efficient problem-solving and data-driven decision-making. Let us understand the steps involved with the help of a flowchart.



IMPLEMENTATION

Let us dive deep into the project & have a look at the queries, data model & all the views created.

Analyzing the startup's table :

Q.1) Calculate the total number of companies in the table.

Query 1.

```
SELECT COUNT (*)
FROM startups;
```

Query Results
COUNT (*)
70

Q.2) Find the total value of all companies.

Query 2.

```
SELECT SUM(valuation)
FROM startups;
```

Query Results
SUM(valuation)
974455790000

Q.3)What is the highest amount raised by a startup?

Query 3.

```
10
11    SELECT MAX(raised)
12    FROM startups;
```

Query Results
MAX(raised)
11500000000

Q.4)What is the highest amount raised by a startup during the ‘Seed’ stage?

Query 4.

```
11    SELECT MAX(raised)
12    FROM startups
13    WHERE stage = 'Seed';
14
```

Query Results
MAX(raised)
1800000

Q.5) In what year was the oldest company on the list founded?

Query 5.

```
15    SELECT MIN(founded)
16    FROM startups;
17
```

Query Results
MIN(founded)
1994

Q.6) What is the average valuation?

Query 6.

```
18  SELECT category, AVG(valuation)
19  FROM startups
20  GROUP BY category ;
21
```

Query Results	
category	**AVG(valuation)**
Ø	4290000.0
Algorithms	7600000.0
Augmented Reality	8000000000.0
Big Data Analytics	15000000.0
Cloud Computing	95000000.0
Customer Service	640000000.0
Data Analytics	Ø
E-commerce	60250000.0
Education	2023800000.0

Q.7) Find the name of each category with the total number of companies that belong to it.

Query 7.

```
27  SELECT category, COUNT(*)
28  FROM startups
29  GROUP BY category;
```

Query Results	
category	**COUNT(*)**
Ø	3
Algorithms	1
Augmented Reality	1
Big Data Analytics	1
Cloud Computing	2
Customer Service	1

Q.8)What are the most competitive markets?

Query 8.

```
30    SELECT category, COUNT(*)
31    FROM startups
32    GROUP BY category
33    HAVING COUNT(*) > 3
34    ORDER BY 2 DESC;
35
```

Query Results	
category	COUNT(*)
Social	12
Mobile	10
Education	5

Q.9) Find the CEO of the highest-value startup.

Query 9.

```
41    SELECT ceo, name, info, MAX(valuation)
42    FROM startups;
```

Query Results			
ceo	name	info	MAX(valuation)
Owen Pilsner	XNA	Genome Database	760000000000

Q.10) What is the average size of a startup in each location?

Query 10.

```
36    SELECT location, AVG(employees)
37    FROM startups
38    GROUP BY location ;
```

Query Results	
location	**AVG(employees)**
Atlanta	3.0
Boulder	3.0
Brooklyn	502.666666666667
Chicago	12.0
Columbus	2.0
Denver	12.0
Fort Lauderdale	500.0
Irvine	2.0
Long Island	5.0
Los Angeles	6.83333333333333
Minneapolis	20.0
New Delhi	250.0
New York	702.75
Omaha	65.0
Palo Alto	125.833333333333
Paris	30.0
San Francisco	1920.4
Savannah	6.0
Scranton	6.0
Seattle	9.66666666666667
Silicon Valley	1804.6
Virginia Beach	15.0
Washington DC	8.0

Chapter 5
Conclusion

In conclusion, we have successfully completed the assigned task of analyzing the "startups" table from the provided "project.sqlite" file. We began by examining the structure of the table and determining the number of columns. Subsequently, we calculated the total number of companies, and the total value of all companies, and identified the highest amount raised by a startup.

We then refined our queries to specifically target the maximum amount raised during the 'Seed' stage and discover the founding year of the oldest company in the list. Additionally, we delved into sector-wise analysis, returning average valuations both overall and within each category. We concluded this section by ordering the categories from highest to lowest averages.

Through these queries and analyses, we have gained valuable insights into the portfolio of major industry players, their financial standings, market competitiveness, and the geographical distribution of startups, all of which can inform strategic decision-making and further research.